UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/693,403	10/24/2003	James R. Stinger	1099.2509-2	1929

7590 06/19/2008

HEWLETT-PACKARD COMPANY
Intellectual Property Administration
P. O. Box 272400
Fort Collins, CO 80527-2400

EXAMINER	
COLAN, GIOVANNA B	

ART UNIT	PAPER NUMBER
2162	

MAIL DATE	DELIVERY MODE
06/19/2008	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Application Number: 10/693,403
Filing Date: October 24, 2003
Appellant(s): STINGER, JAMES R.

LeRoy D. Maunu
For Appellant

EXAMINER'S ANSWER

This is in response to the appeal brief filed 03/24/2008 appealing from the Office action mailed 10/18/2007.

(1) Real Party in Interest

A statement identifying by name the real party in interest is contained in the brief.

(2) Related Appeals and Interferences

The examiner is not aware of any related appeals, interferences, or judicial proceedings which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

(3) Status of Claims

The statement of the status of claims contained in the brief is correct.

(4) Status of Amendments After Final

The appellant's statement of the status of amendments after final rejection contained in the brief is correct.

(5) Summary of Claimed Subject Matter

The summary of claimed subject matter contained in the brief is correct.

(6) Grounds of Rejection to be Reviewed on Appeal

The appellant's statement of the grounds of rejection to be reviewed on appeal is correct.

(7) Claims Appendix

The copy of the appealed claims contained in the Appendix to the brief is correct.

(8) Evidence Relied Upon

6,336,124	ALAM	10-1998

(9) Grounds of Rejection

The following ground(s) of rejection are applicable to the appealed claims:

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

Claims 1, 3, 5, 7, 9, 12, and 15 – 20 are rejected under 35 U.S.C. 102(e) as being anticipated by Alam et al. (Alam hereinafter) (US Patent No. 6,336,124 B1, filed: October 1, 1998).

Regarding Claim 1 and 7, Alam discloses a computer-readable medium having stored thereon sequences of instructions, said sequences of instructions including

instructions which, when executed by a processor, cause said processor to perform the steps of:

receiving a page description language representation of the document (Col. 6, lines 24 – 28, Alam) for providing a list of words in the document (Col. 6, lines 57 – 59, words, Alam) and position information for the words (Col. 6 and 7, lines 54 – 57 and 60 – 63; respectively, X and Y coordinates, Alam); and

automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature (Col. 7, lines 6 – 9, 32 – 35, and 49 – 55, "an element of a table, or all portion of the table, depending upon the spacing of the elements of the table…", and "results from subsequent steps, such as steps 702, 704, 706, and 708, are used to obtain the information which would otherwise be contained in the tags. The subsequent steps utilize the layout information (i.e., image representation) of the text to of the document locate words, lines, paragraphs, and tables, for example….", respectively, Alam), wherein the identifying step includes:

dividing the document into one or more pages (Col. 15, lines 56 – 58, dividing into sub-pages, Alam);

dividing each page into a plurality of lines (Col. 15, lines 57 – 60, Alam); for each line (Col. 8, lines 57 – 60, Alam), clustering the words of the line into one or more word clusters (Col. 9, lines 1 – 6, Alam[1]), wherein each cluster includes one or more words (Col. 16, lines 15 – 22; if the **block is a paragraph that ends**

with an incomplete sentence or an improper termination … whether **the block contains one or more sentences**, Alam), each cluster having a horizontal beginning point, horizontal midpoint, and horizontal end point (Col. 11, lines 8 – 17, wherein the right most X coordinate corresponds to the horizontal beginning point claimed; wherein the center X coordinate corresponds to the horizontal midpoint claimed; and wherein the left X coordinate corresponds to the horizontal end point claimed; Alam);

for clusters in the plurality of lines, comparing alignment of the horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines, wherein a cluster in a first line is aligned with a cluster in a previous line if at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the first line is aligned with at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the previous line (Col. 10, and 11, lines 24 – 33, and 4 – 7, "…To determine whether the selected line is near the current paragraph in the Y direction, the appropriate Y coordinate (s) of the selected line are compared with the appropriate Y coordinate(s) of the previous line of the current paragraph to determine whether certain parameters and/or thresholds are satisfied…"; and "… comparing the left X coordinate of the first word of the current line with the left X coordinate of the first word of a first line in the current paragraph to account for the hanging indent…"; respectively, Alam); and

[1] Wherein examiner interprets the step of determining the spacing between words as the step of

identifying a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table (Col. 12, and 17, 47 – 49 and 51 – 52, and 10 – 18; "...determines if the current block is not a table, step 1904 breaks up the current block into elements such that ach element can be displayed within the display configuration. Each element of a paragraph block may be, for example, a word contained in the paragraph. Other division of a block into elements may be implemented. For example, each element of a list block may be an item or a line in the list..."; respectively, Alam); and

outputting data descriptive of the lines of the table (Col. 1 – 2, and 19, lines 65 – 67 and 1, and 48 – 60; "...at least a portion of the first row forming the row heading of sample table 2200 is displayed in each of the display pages...", Alam).

Regarding Claim 3, Alam discloses a method wherein the step of automatically identifying table data in the document based on the number of word clusters for each line and the alignment of the word clusters between lines (Col. 12, lines 47 – 49 and 51 – 52, Alam) further comprises:

using the word clusters to generate column position information (Col. 9 and 18, lines 10 – 12 and 9 – 12; respectively, Alam), wherein the column information includes for each column a horizontal beginning point, horizontal midpoint, and horizontal end point (Col. 11, lines 8 – 17, wherein the right most X coordinate corresponds to the

---

clustering the words claimed.

horizontal beginning point claimed; wherein the center X coordinate corresponds to the horizontal midpoint claimed; and wherein the left X coordinate corresponds to the horizontal end point claimed; Alam); and

updating the column position information (Col. 10, lines 42 – 45, Alam[2]) by performing a union operation between the column position information of a previous line (Col. 10, lines 26 – 29, Y coordinates of the previous lines, Alam) and the column position information of a current line (Col. 10, lines 24 – 27, Y coordinates of the selected lines, Alam).

Regarding Claim 5, Alam discloses a method wherein receiving a page description language representation of the document for providing a list of words in the document and position information for the words includes receiving a PDF representation of the document (Col. 2, lines 1 – 5, portable document format (PDF), Alam), and wherein converting the table data encompassed by each table bounding box to a markup language representation includes converting the table data encompassed by each table bounding box to a HTML representation (Col. 6, lines 42 – 47, Alam).

Regarding Claim 9, Alam discloses a computer-readable medium further containing instructions which, when executed by said processor, would cause said processor to perform the steps of:

[2] Wherein examiner interprets the step of determining that the line is not near the current paragraph as the step of updating position information claimed.

using the word clusters to generate column position information (Col. 9 and 18, lines 10 – 12 and 9 – 12; respectively, Alam), wherein the column information includes for each column a horizontal beginning point, horizontal midpoint, and horizontal end point (Col. 11, lines 8 – 17, wherein the right most X coordinate corresponds to the horizontal beginning point claimed; wherein the center X coordinate corresponds to the horizontal midpoint claimed; and wherein the left X coordinate corresponds to the horizontal end point claimed; Alam); and

updating the column position information (Col. 10, lines 42 – 45, Alam[3]) by performing a union operation between the column position information of a previous line (Col. 10, lines 26 – 29, Y coordinates of the previous lines, Alam) and the column position information of a current line (Col. 10, lines 24 – 27, Y coordinates of the selected lines, Alam).

Regarding Claim 12, Alam discloses a document processing system comprising:

a processor for executing programs (Fig. 2, item 151, Col. 4, lines 61 – 64, Alam); and

a table identification program (Col. 4, lines 43 – 46, Alam) for receiving a page description language representation of a document (Col. 6, lines 24 – 28, Alam), the page description language representation providing a list of words in the document (Col. 6, lines 57 – 59, words, Alam) and position information for the words (Col. 6 and 7, lines

[3] Wherein examiner interprets the step of determining that the line is not near the current paragraph as the step of updating position information claimed.

54 – 57 and 60 – 63; respectively, X and Y coordinates, Alam), and for automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature (Col. 7, lines 6 – 9, 32 – 35, and 49 – 55, "an element of a table, or all portion of the table, depending upon the spacing of the elements of the table…", and "results from subsequent steps, such as steps 702, 704, 706, and 708, are used to obtain the information which would otherwise be contained in the tags. The subsequent steps utilize the layout information (i.e., image representation) of the text to of the document locate words, lines, paragraphs, and tables, for example….", respectively, Alam), wherein the identification program is configured to,

dividing the document into one or more pages (Col. 15, lines 56 – 58, dividing into sub-pages, Alam);

dividing each page into a plurality of lines (Col. 15, lines 57 – 60, Alam); for each line (Col. 8, lines 57 – 60, Alam), cluster the words of the line into one or more word clusters (Col. 9, lines 1 – 6, Alam[4]), wherein each cluster includes one or more words (Col. 16, lines 15 – 22; if the **block is a paragraph that ends with an incomplete sentence** or an improper termination … whether **the block contains one or more sentences**, Alam), each cluster having a horizontal beginning point, horizontal midpoint, and horizontal end point (Col. 11, lines 8 – 17, wherein the right most X coordinate corresponds to the horizontal beginning point claimed; wherein the center X coordinate corresponds to the horizontal

midpoint claimed; and wherein the left X coordinate corresponds to the horizontal end point claimed; Alam);

for clusters in the plurality of lines, compare alignment of the horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines, wherein a cluster in a first line is aligned with a cluster in a previous line if at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the first line is aligned with at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the previous line (Col. 10, and 11, lines 24 – 33, and 4 – 7, "…To determine whether the selected line is near the current paragraph in the Y direction, the appropriate Y coordinate (s) of the selected line are compared with the appropriate Y coordinate(s) of the previous line of the current paragraph to determine whether certain parameters and/or thresholds are satisfied…"; and "… comparing the left X coordinate of the first word of the current line with the left X coordinate of the first word of a first line in the current paragraph to account for the hanging indent…"; respectively, Alam); and

identify a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table (Col. 12, and 17, 47 – 49 and 51 – 52, and 10 – 18; "…determines if the current block is not a table, step 1904 breaks up the current block into elements such that ach element can be displayed within the display configuration. Each

[4] Wherein examiner interprets the step of determining the spacing between words as the step of

element of a paragraph block may be, for example, a word contained in the paragraph. Other division of a block into elements may be implemented. For example, each element of a list block may be an item or a line in the list…"; respectively, Alam); and

output data descriptive of the lines of the table (Col. 1 – 2, and 19, lines 65 – 67 and 1, and 48 – 60; "…at least a portion of the first row forming the row heading of sample table 2200 is displayed in each of the display pages…", Alam).

Regarding Claim 15, Alam discloses a document processing system of claim 13 wherein the table identification program further comprises:

a conversion module (Fig. 6, item 628, Col. 6, lines 41 – 45, converter, Alam) coupled to the bounding box generation module for receiving the table bounding box for each table in the document (Fig. 6, item 612, Col. 6, lines 35 – 40, ACROBAT CAPTURE and ACROBAT WRITER, Alam), and for converting the words encompassed by the table bounding box into a markup language representation that maintains the table structure of each table (Col. 6, lines 42 – 47, Alam).

Regarding Claim 16, Alam discloses a method wherein the step of automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature further comprises:

---

clustering the words claimed.

automatically identifying table data in the document based on one or more table headings (Col. 2, lines 39 – 44, Alam).

Regarding Claim 17, Alam discloses a method wherein the step of automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature further comprises:

automatically identifying table data in the document based on one or more horizontal lines (Fig 15B, Col. 14, lines 38 – 43, five rows, Alam) and vertical lines that separate rows or columns of the table (Fig. 15C, Col. 14, lines 46 – 49, two columns, Alam).

Regarding Claim 18, Alam discloses a method, wherein the step of automatically identifying table data in the document based on the number of word clusters for each line and the alignment of the word clusters comprises:

determining whether the number of word clusters in a line is greater than a threshold value (Col. 8, lines 21 – 26; If the inter-word spacing or distance in the Y direction is greater than a threshold…; Alam); and

classifying the word clusters in the line as a row of a table in response to the number of word clusters in a line being greater than the threshold value (Col. 8, lines 28 – 34; Alam).

Regarding Claim 19, Alam discloses a computer-readable medium, wherein the instructions for automatically identifying table data in the document based on the number of word clusters for each line and the alignment of the word clusters include instructions that when executed by a processor cause the processor to perform the steps further comprising:

determining whether the number of word clusters in a line is greater than a threshold value (Col. 8, lines 21 – 26; If the inter-word spacing or distance in the Y direction is greater than a threshold…; Alam); and

classifying the word clusters in the line as a row of a table in response to the number of word clusters in a line being greater than the threshold value (Col. 8, lines 28 – 34; Alam).

Regarding Claim 20, Alam discloses a document processing system, wherein the table identification program is further configured to:

determine whether the number of word clusters in a line is greater than a threshold value (Col. 8, lines 21 – 26; If the inter-word spacing or distance in the Y direction is greater than a threshold…; Alam); and

classify the word clusters in the line as a row of a table in response to the number of word clusters in a line being greater than the threshold value (Col. 8, lines 28 – 34; Alam).

(10) Response to Argument

Claims 1,3, 5, 7, 9, 12, and 15-20 stand rejected under 35 U.S.C. §102(e) as being anticipated by "Alam" (US Patent 6,336,124 to Alam et al.).

Claims 1, 5, 7, 12, and 16

In response to appellant's argument that the references fail to show certain features of appellant's invention, it is noted that the features upon which appellant relies (i.e., "no teaching here of 'how each line' is identified as being part of a table..." in page 6, and "comparison of any horizontal midpoint of word..." in page 9; appeal brief) are not recited in the rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

Appellant argues that; "the Examiner has not shown that Alam teaches the claimed "identifying a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table"; and further that the teachings of Alam; "that the Examiner cited as corresponding to the claimed 'comparing alignment of the horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines' do not relate to the identifying a line as being part of a table".

Examiner respectfully disagrees. Alam does disclose: identify a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table (Col. 12, and 17, 47 – 49 and 51 – 52, and 10 – 18; "...determines if the current block is not a table, step 1904 breaks up the current block into elements such that each element can be displayed within the display configuration. Each element of a paragraph block may be, for example, a word contained in the paragraph. Other division of a block into elements may be implemented. For example, each element of a list block may be an item or a line in the list..."; respectively, Alam). Alam also discloses: comparing (Col. 11, lines 5 – 8, "... comparing the left X coordinate of the first word of the current line with the left X coordinate of the first word of a first line in the current paragraph to account for the hanging indent..."; Alam) alignment of the horizontal beginning point (Page 11, 8 – 17, "If the right X coordinate of the last word of the current line is within a threshold distance from the right-most X coordinate of the last words of the

lines of the current paragraph…", Alam), horizontal midpoint (Page 11, 8 – 17, "… if the center X coordinate of the current line,...", Alam), and horizontal end point of clusters between lines (Page 11, 8 – 17, "…the average of the left X coordinate of the first word and the right X coordinate of the last word of the current line, is within a threshold distance less or greater than the center X coordinate of the previous existing line in the current paragraph…", Alam), wherein a cluster in a first line is aligned with a cluster in a previous line if at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the first line is aligned with at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the previous line (Col. 11, lines 8 – 25, "…and the current paragraph may be right aligned and thus overlap. Further, if the center X coordinate of the current line, i.e., the average of the left X coordinate of the first word and the right X coordinate of the last word of the current line, is within a threshold distance less or greater than the center X coordinate of the previous existing line in the current paragraph, i.e., the average of the left X coordinate of the first word and the right X coordinate of the last word of the previous existing line of the current paragraph, then both the selected line and the current paragraph may be center aligned and thus overlap. The threshold distance may be, for example, 0.5 of the width of a character of the average width of a character. The above are merely illustrative examples for determining the alignment of the lines and whether a line near a paragraph is similarly aligned…."; Alam). It is clear from these paragraphs that Alam discloses a "comparing of alignment of the horizontal beginning point, horizontal point …" to determine "…if at

least one…is aligned…". According to Alam, after determining "if aligned", the lines are joined into paragraphs (see Fig. 10, item 1010 "Does the selected line overlap the current paragraph vertically (Left/right/center aligned)", 1012, 1016 "Add selected line to current paragraph", and 1018, Col. 11, lines 53 – 55, "If all lines have been assigned to a paragraph, the process of joining lines into paragraph is complete", Alam) and further identified as being part of a table (See for example; Col. 12, lines 35 – 44, "After the words are joined into lines and the lines joined into paragraphs, tables are located… Any suitable method may be utilized to locate tables from the joined paragraphs…Alam, discloses a processor based method for recognizing, capturing…tabular data…", Alam; and for more details see: Col. 17, lines 10 – 18, "step 1902 determines if the current block is a table. If the current block is not a table, step 1904 breaks up the current block into elements such that each element can be displayed within the display configuration. Each element of a paragraph block may be, for example, a word contained in the paragraph. Other division of a block into elements may be implemented. For example, each element of a list block may be an item or a line in the list…", lines 24 – 31, "Each element of the current block may comprise a word or a line, for example, which can be broken up into multiple lines and/or multiple words. If the current block is a table, the first row and first column of the table are selected as the row and column headings at step 1905. Although not all first rows and first columns of tables are headings, it can be assumed that the first row and first column are headings…"; wherein the Examiner interprets the step of determining if a block is a table, wherein Alam's block includes lines, as the step of identifying a line

as being part of a table as claimed; Alam). Finally, the flow of events of Alam's disclosure indicate that the step of comparing alignments is for the purpose of "identifying a line as being part of a table" as claimed.

Claims 3 and 9

Appellant's argument that "…there is no apparent teaching or suggestion by Alam that the word clusters are used to generate column information that includes a horizontal midpoint..." has been fully considered but it is not persuasive.

Alam does disclose: using the word clusters to generate column position information (Col. 9 and 18, lines 10 – 12 and 9 – 12; respectively, and Page 17, lines 25 – 40, "Each element of the current **block may comprise a word** or a line, for example, which can be broken up into multiple lines and/or multiple words. If the current block is a table, the first row and first column of the table are selected as the row and column headings at step 1905… Step 1906 **determines the number of columns n that can be displayed with the column heading**, if any, within the display configuration. The n non-heading columns are then selected and the selected elements or columns of the first row are added to a subblock set as the current sub-block at step 1907…," Alam), wherein the column information includes for each column a horizontal beginning point, horizontal midpoint, and horizontal end point (Examiner makes note that Alam discloses that the blocks include words, and further that words include the coordinates disclose in Col. 11, lines 8 – 17, wherein the right most X coordinate corresponds to the horizontal

beginning point claimed; wherein the center X coordinate corresponds to the horizontal midpoint claimed; and wherein the left X coordinate corresponds to the horizontal end point claimed; Alam).

Claims 18, 19, and 20

Appellant argues that; "Alam does not determine whether the number of word clusters in a line is greater than a threshold value".

Examiner respectfully disagrees. Alam does disclose: determining whether the number of word clusters in a line is greater than a threshold value (Col. 8, lines 21 – 26; If the inter-word spacing or distance in the Y direction is greater than a threshold…; Alam).

In response to appellant's argument that the references fail to show certain features of appellant's invention, it is noted that the features upon which appellant relies (i.e., considers the number of words in a line) are not recited in the rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

Appellant argues that; "that Alam does not classify the word clusters in the line as a row of a table in response to the number of word clusters in a line being greater than the threshold value".

Examiner respectfully disagrees. Alam does disclose: classifying the word clusters in the line as a row of a table in response to the number of word clusters in a

line being greater than the threshold value (Col. 8, lines 28 – 34; and Col. 18, lines 5 – 9, Alam).

(11) Related Proceeding(s) Appendix

No decision rendered by a court or the Board is identified by the examiner in the Related Appeals and Interferences section of this examiner's answer.

For the above reasons, it is believed that the rejections should be sustained.

Respectfully submitted,

/Giovanna Colan/

Examiner, Art Unit 2162

Conferees:

/John Breene/

Supervisory Patent Examiner, Art Unit 2162

/Eddie C Lee/

Supervisory Patent Examiner, TC 2100

An appeal conference was held on 02 June 2008, and it was agreed to proceed to the board of appeals.